DEPARTMENT OF HEALTH & HUMAN SERVICES Public Health Service

Memorandum

Food and Drug Administration
Center for Biologics Evaluation and Research
1401 Rockville Pike
Rockville, MD 20852

Division of Clinical Trial Design and Analysis
HFM-576

Date: December 15, 2000

From: Marc Walton, MD, PhD; OTRR/DCTDA

Subject: Medical Officer Review

Through: Karen Goldenthal, MD, Director, OVRR/DVRPA

PLA 91 – 0184
STN 103000 / 1004
Second CR Letter Response Submission

Allergan, Inc.

BOTOX
Botulinum Toxin Type A

For Treatment of Cervical Dystonia

Clinical Review

OVERVIEW

In June 1999 Allergan, Inc. submitted a complete response to a December 1995 Complete Review Letter for a supplemental marketing application, PLA 91-0184 for botulinum toxin type A neurotoxin complex for use in the treatment of cervical dystonia. The June 1999 submission has been fully reviewed and is the subject of a complete review document, dated November 19, 1999.

The initial submission of PLA 91-0184 occurred in March 1991, with an amendment submission in March 1994. The December 1995 Complete Review Letter to Allergan stated the submitted information was inadequate to provide marketing approval for the cervical dystonia indication, and an additional phase 3 clinical study was required. Allergan conducted such a study, and submitted it as the June 1999 submission, along with additional studies determined to be necessary during IND discussions. Allergan has proposed the supplemental indication be stated in the labeling as: Botox is indicated for the treatment of cervical dystonia (spasmodic torticollis) in adults.

History of the sPLA and Scope of this Review

The initial studies of Botox in the treatment of cervical dystonia which were the basis of the 1991 sPLA submission were deemed to be seriously flawed in design and conduct and inadequately documented. There were 5 controlled trials and 3 open label treatment studies in this initial group. A brief orientation to those 5 studies is contained in an appendix of the November 1999 review document. For detailed review information the comprehensive reviews of the 1991 sPLA by Dr. L. Teague (CBER) and Dr. Collins (CDER Collaborative Review) should be consulted, as well as the CBER Statistical Review of those studies.

There were four studies presented in the June 1999 submission. Study 140 is the phase 3 efficacy trial, and Study 147 is a companion study to assess the evaluation tool employed in Study 140. Subsequent to completing Study 140, Allergan changed the marketed toxin source from the 1979 single batch (the only source of Botox in the U.S. since the initial marketing approval) to a batch produced in a newly constructed manufacturing facility. This new toxin, designated [redacted], received marketing approval in late 1997 for the labeled indications. However, the studies evaluating [redacted] raised concern regarding an increased risk of regional toxin spread. Allergan was asked to address this potential difference in safety profile with additional studies for the CD indication. Two additional studies, Study 004 and 014 were submitted to address the safety of the [redacted] Botox product in the CD indication. Full details of study design and results of these four studies, along with background information regarding the disease, product class, and overall clinical development program were the subject of the November 1999 Clinical Review. That review document should be consulted for extensive information regarding those topics, and will not be repeated in this review document; being assumed to be available and/or known to the reader of this review document.

That review culminated in a Complete Review Letter, dated December 9, 1999, in which further information was requested regarding certain issues raised by the review, and not adequately addressed in the Allergan submission. This review will be limited to the information submitted by Allergan in response to the December 1999 CR letter, and the overall sPLA recommendation. The Allergan documents reviewed include the February 4, 2000 submission (the majority of the information) as well as small additional submissions March 7, 2000, August 25, 2000, September 27, 2000, October 4, 2000 and several largely labeling submissions subsequently.

TABLE OF CONTENTS

ADDITIONAL SUMMARY OF ADMINISTRATION METHOD

A summary of dose administration regarding frequency of usage of the individual muscles, and of the dose per muscle had been already submitted. However, the summary on the number of muscles injected per subject had not been. Allergan responded to the request with a summary of Period I, and the toxin treated group in Period II.

Table 1: Number of Muscles Used in Each Subject in Treatment Session			
Number of Muscles	Period I	Period II Toxin Group	Combined
2	13%	6%	11%
3	37%	43%	39%
4	31%	32%	31%
5	13%	17%	14%

Thus, most subjects had 3 or 4 muscles injected in a treatment session, with much fewer receiving injections to 2 or 5 muscles.

Allergan has also acknowledged that dose administration information is appropriate for labeling, and proposed that typical dose by muscle be shown, employing the 25th to 75th percentiles as the indication of typical dose.

Table 2: Range of Dose by Individiual Muscle in Study 140 - BOTOX Subjects of Period II				
Muscle	n	Mean (U)	25th Percentile	75th Percentile
All combined	88	236	198	300
Sternocleidomastoid	77	55	40	70
Trapezius	49	70	25	100
Levator Scapulae	52	49	25	60
Splenius capitis	83	87	60	100
Scalene (any)	15	42	15	54
Other muscles	34	71	30	100

Comment:

This appears to be a reasonable presentation of the summary data on dose administration. The information that approximately 70% of subjects received injections to 3 or 4 muscles will be important to add to the proposed labeling information. However, the numbers of subjects in this table is relatively few, especially for all but the most frequently used muscles. Examination of these doses by muscle for the Period I treatment (with more than twice as many treatments) would be important to ensure that the typical dosing range is not being misrepresented. Additionally, the dose per muscle should be examined as percent of total treatment dose; if this is more narrow in range, than it may be a better presentation of the data.

In response to request from CBER, Allergan has also submitted the dosing information in the form of dosing to each muscle as percent of total dose delivered at the treatment session.

Table 3: Range of Dose to Individiual Muscle in Study 140 - BOTOX Subjects Only - Dose as Percent of Session Total Dose								
	Period II				Period I			
Muscle	n	25th Percentile	Median	75th Percentile	n	25th Percentile	Median	75th Percentile
All combined	88	198 U	236 U	300 U	214	200 U	240 U	300 U
Sternocleidomastoid	77	17	23	31	184	19	25	30
Trapezius	49	18	26	33	118	17	25	33
Levator Scapulae	52	16	20	25	123	12	18	27
Splenius capitis	83	25	34	50	207	28	38	50
Scalene (any)	15	6	17	21	36	8	15	20
Other muscles	34	14	25	38	72	15	25	35

Comment:

This appears to be an improved manner of presentation of the information. The apparent variation in dose in each muscle is more limited than in the table presenting dose in units. The dose information from Period I is consistent with the information from Period II. This is the form that should be employed in product labeling.

Allergan subsequently further considered the issue of how to provide guidance to physicians employing BOTOX for this indication, and performed analyses of dosing within each muscle for groups of subjects with either the same 3 or 4 muscles injected for the most common groupings of 3 or 4 muscles.

Table 4: Dosing in Toxin Units to Muscles in Cohorts of Same Same Muscle Use																			
	Cohort A n = 45			Cohort B n = 38			Cohort C n = 21			Cohort D n = 19			Cohort E n = 18			All Pd II Subj n = 87			
Muscle	25Ptl	Mdn	75Ptl	25Ptl	Mdn	75Ptl	25Ptl	Mdn	75Ptl	25Ptl	Mdn	75Ptl	25Ptl	Mdn	75Ptl	n	25Ptl	Mdn	75Ptl
Spelnius Capitus	60	90	100	50	75	100				60	75	100	60	70	100	83	60	90	100
Sternocleidomastoid	45	50	70	40	50	65	40	50	65				40	50	65	77	40	50	70
Levator Scapulae	25	50	60				25	35	60	25	40	60	25	38	60	52	25	50	60
Trapezius				32	60	100	30	60	100	25	60	100	25	50	100	49	35	60	100
All focus Muscles	150	185	225	150	195	265	110	155	300	120	175	225	180	227	285	87	170	200	285

Comment:

This division of patients into cohorts seems to add little to what was already known. The range and median in Units for each of the identified muscles in each of the specialized same-subject cohorts is very little different from the last grouping of data, which is all subjects. This suggests in fact that the dosing to each of these muscles was not varied in relation to the other specific muscles being injected. These specialized cohorts would not add anything useful to the labeling.

STUDY 140 QUALITY OF STUDY CONDUCT TOPICS

VISIT SCHEDULE TOLERANCE COMPLIANCE

Allergan states that for Period I there were 7 subjects of the 214 with visits outside of the prescribed visit dates. Only two of these were for the Week 6 visit, the most important. Both if these were less than 2 weeks distant from the ideal date, but greater than the protocol specified ±1 week limit.

For Period II there were 6 subjects with visits outside of the defined time window, again only two for the Week 6 evaluation. These 2 subjects had evaluations that were 3 and 4 days later than the within 1-week limits. An analysis by Allergan has been provided with treatment of these as missing values.

Comment:

These errors were few, and relatively minor in importance. These should not have any significant impact on the study results and can be ignored in analyzing the data.

SUBJECTS WITH TWO PERIOD I ENROLLMENTS OR ERRONEOUS PERIOD I RETREATMENT

Allergan reports that 7 subjects had more than 1 enrollment into Period I; one of these subjects in fact had 3 enrollments into Period I. Of these, 5 were for Period I eligibility violations not initially recognized of requiring 2 prior BOTOX treatment sessions 12 – 16 weeks apart, and within 12 – 16 weeks of study enrollment. Apparently the erroneous enrollment into Period I lead to meeting this eligibility requirement, so that at the end of the Period I follow-up, the subjects were validly eligible for enrollment, and so were re-enrolled into Period I. They received new Period I subject numbers at that time. Of these 5, 4 subjects then went on to enrollment into Period II (2 placebo, 2 BOTOX). The fifth of these subjects (#007) was erroneously retreated with a Period I drug supply for a third treatment session, and this subject was followed, but did not advance into Period II.

Two other subjects (Period I # 159, 607) were erroneously re-injected with a Period I drug supply when they were intended to advance to Period II; neither of these subjects subsequently advanced into actual Period II randomization either. Both of these errors were quickly discerned by the site PI. Only the first Period I enrollment of each of these was used in Period I analyses.

With regards to these last two subjects, Allergan notes that the failure to properly enroll subject 159 into Period II altered the randomization assignments for 7 subjects; Subject 159 who was not enrolled, and 6 of the 9 subjects at the site who were enrolled into the study after the error. There were no further site-specific consequences after subject 607, as no further subjects were enrolled at that site. Allergan reports there were 3 subjects who were enrolled at the site after Subject 007, all 3 would have been assigned to the opposite treatment group had the error not occurred.

Note that Allergan subsequently incorporates these effects into many of the sensitivity analyses submitted later in their materials.

Comment:

These errors appear to have been spread out at different sites and do not reasonably appear to be intentional. The ability to have intentionally altered assignments for multiple patients appears to have been absent, as the subsequent enrollment of subjects either did not occur or occurred so long after the error that it would have been quite unlikely to be able to predict

which would be the next subject to be randomized at that site. Thus, there does not appear to have been any attempt to circumvent the proper randomization process, and the assignments of the subsequent subjects at the sites can be accepted as unbiased. The missing 3 subjects from the study is a fairly minimal number, and unlikely to have altered the study outcome substantially, irrespective which group they were randomized to or whether or not they responded. Therefore, the study does not appear to have been harmed from these relatively few (7 of over 200 subjects in Period I) errors of study conduct. Sensitivity analyses should be performed without consideration of these errors, and should not employ the non-enrolled subjects.

LACK OF EFFICACY DISCONTINUATIONS

Allergan reports that these occurred by subject choice, based on the subject's opinion regarding the benefit received in the study. The study follow-up ended at the time the subject informed the Investigator of the perceived lack of efficacy and wish to discontinue; if this was prior to Week 6, then this would result in a Week 6 missing value. If this was at the Week 6 visit, then the exam would occur, and no Week 6 missing value occurred; although values were missing for subsequent visits.

Allergan reports that there were only 7 Lack of Efficacy discontinuations resulting in a missing Week 6 value during Period II. Of these 7, 3 were in the placebo group, 4 in the Botox group. Most Lack of Efficacy determinations and discontinuation occurred at the time of the Week 6 visit (14), with 6 LoE discontinuations at Week 2 or 4, and 6 more at Week 8. Only the 6 at Week 2 or 4 resulted in LoE missing values. There was one additional LoE discontinuation subject with missing Week 6 evaluation; however this subject, assigned to the Botox group, had an unscheduled visit 1 week after the planned Week 6 visit, and just within the protocol defined timewindow for visit compliance. The subject declared the intent to discontinue for lack of efficacy at that time. For this subject, Allergan has used the unscheduled visit evaluation in place of treating this as a true missing value, and employing the imputation rule. Therefore, Allergan used the LoE missing rule for only 6 subjects total in Period II.

Comment:

There were fairly few Lack of Efficacy Discontinuations that resulted in missing values for the primary endpoint. Most LoE subjects were able to participate until the Week 6 evaluation of Period II. It is reasonable to employ the unscheduled visit information for the one BOTOX group subject (#403) who missed the planned Week 6 visit, but did present in clinic within the specified time-window rather than make that subject a missing value for Week 6. This subject is more like the LoE discontinuations who declared LoE at the week 6 visit, and only visits subsequent are actually missing.

PERFORMANCE OF OUTCOME ASSESSMENTS

Allergan reports that the protocol did require the PI to perform all outcome evaluations, and no record was kept of who performed each evaluation for any specific subject. Allergan acknowledges that improvements have been made in current studies, and ongoing studies do document who is authorized to perform specific evaluations. However, the protocol did direct that the primary endpoints of CDSS and Global Assessment be performed by the same evaluator within each subject's participation. There was no documentation as to degree of compliance with that directive.

Comment:

This is an unfortunate lapse in the record keeping and CRF documentation for this study. However, no other issues of biased study conduct have been discerned in review or site inspections to support any concern of intentional bias. If the study remained well blinded then any failure to maintain consistency should be unbiased between the treatment groups, and not damage the acceptability of study results.

STUDY 140 EFFICACY ENDPOINT ANALYSES

APPARENT DISCREPANCIES AND ANOMALIES IN SUBMITTED DATASETS AND ANALYSES

The analytic deviations noted in the original CBER review included:

Use of different LoE imputation for missing values for the treatment groups

Subject 653 missing Period II baseline used Period I baseline CDSS imputation, and dropped from Global Assessment analysis.

Subject 403 reported to be a LoE subject, and missing Week 6 evaluation, but not treated per analytic plan as an LoE subject

Use of an LoE imputation for Subject 552 for Global Assessment who was not a LoE discontinuation.

Comment:

Although these are few in number, they become critical in a study as small as this one, and when the outcome is of marginally significant (if at all) treatment effects.

In addition to the issue of method of LoE imputation, Allergan discusses all the study subjects raised in the CR letter to Allergan. With regard to the imputation of missing value for LoE which was different by treatment group, Allergan acknowledges that the written analytic plan did not provide for this, but states that the Allergan statistician had assumed that differential imputation would be implicitly clear to readers of the plan.

Comment:

This suggestion carries no substance. The purpose of written, prospective plans is to state clearly, and for all readers, exactly what approach is planned. CBER at no time was informed of the plan to impute differently between the treatment groups, nor was the opportunity of numerous phone discussions in which the Allergan statistician participated employed to convey this information. CBER views the differential imputation to be a purely post hoc modification instituted after unblinding of the study.

Allergan states that the imputation of worst value over all study subjects is unduly harsh due to selecting the most extreme outlier value. Allergan prefers alternate methods, such as using the average value among all the non-missing observations as the imputed value. Allergan notes this approach to missing values has been requested by FDA for other Allergan studies.

Comment:

Allergan's concern that this imputation method emphasizes the most extreme outlier is a valid concern. However, it was the method selected by Allergan, and not one proposed by CBER; CBER merely allowed it as a not unreasonable choice while requesting that sensitivity analyses also be performed to assess the consistency of this method. It remains the method that was prospectively planned; all other methods are post hoc selections of a primary analysis method. As to other methods being requested by CBER in other studies; while this is correct, the circumstances of these other studies, being done in markedly

different indications, are different. An imputation method most suitable for one study in one indication with one specific assessment tool may well not be the most appropriate for a different study in a different indication with a different assessment tool. Nonetheless, sensitivity analyses with various imputation methods are appropriate to aid an overall understanding of the meaning of the study results.

With regards to specific subject values, the reason for not using the LoE missing value method for subject #403 (BOTOX) was discussed above, in which the value of improvement by 5 CDSS points is asserted to be the true observation, and not appropriate to impute the worst score. Allergan provides a clarification that the reported imputed value for Subject 459 was a typographical error in the study report, and was correctly handled in the dataset. For Subject 552 (Placebo) Allergan explains that there were no observations for the Physician Global Assessment in Period II, so that LOCF imputation was not feasible for this non-LoE discontinuation. Therefore, they employed the LoE imputation method because it had been a prespecified method, even if not for this circumstance. Allergan does not provide any further justification for their handling of Subject 653 (placebo group) who had a missing baseline assessment for Period II.

Comment:

For Subject 403 Allergan's method appears to be reasonable and the CDSS improvement of 5 points appears to be appropriate to employ even though the subject felt they discontinuing for lack of efficacy. For Subject 552 the attribution of worst observed score for the Global Assessment is not appropriate because this was not an LoE discontinuation. However, for the primary analysis of this endpoint, percentage with any improvement, this should not be important as even attributing a score of 0 (no change) would yield the same percentage with improvement within the group. This anomaly of imputation would bias the tertiary analyses of mean score on this endpoint. Allergan's imputation of baseline for Period II by using the baseline of Period I remains unreasonable given the subject's absence of ever having a score at that level of severity at any other time in the study, and especially at the late observations in Period I or II. LOCF would appear to be a more appropriate method to attribute a baseline for this patient for Period II.

In conclusion, the appropriate prospectively planned analysis was not carried out by Allergan. There were important anomalies, and many of these appear to have occurred in a post hoc manner. Appropriate analyses have been requested from Allergan.

REVISED PRIMARY ENDPOINT ANALYSES AND SENSITIVITY ANALYSES

Revised analyses were requested from Allergan to be ITT (inclusion of all randomized subjects) and with adherence to the analytic plan. There was apparent misunderstanding of exactly how to perform this, and the initial submission included in the analyses the Dual-Period I subjects who were never included in Period II, as well (perhaps) of reassignment of several subjects who would have been randomized differently if those subjects were included. Discussions with Allergan were conducted, and the actual intent of the requests re-clarified. Allergan was also requested to submit CDSS analyses that examined percentage change in score, and that analyses should show observed parameter for each group, estimated treatment effect, confidence interval on the treatment effect, and associated p-value.

The revised analyses were based on the following modifications to the original submission:

LoE imputation would be identical for both treatment groups

Visit time window strict adherence would not be required
No subjects not actually randomized into Period II would be included in the analysis
No changes in the assignment of subjects who might have received a different assignment if those non-randomized subjects had actually enrolled into Period II.
Sensitivity analyses of Mean non-missing value for LoE discontinuations; for all missing values; use of non-parametric analytic method, drop missing value subjects; LOCF all missing values. Also logistic regression for the Global Assessment endpoint rate of responder analysis.
Use of last available CDSS as the baseline for Period II for Subject 653
Use of Unscheduled visit value for Subject 403
Subject 552 missing Global Assessment value imputation not revised.

Table 5: Revised Analyses of Week 6 CDSS Change from Baseline and Sensitivity Analyses

Missing Value Imputation	Placebo n = 82	Botox n = 88	Difference	p-value
Prospective Analytic Plan	-0.27	-1.26	-0.98	0.13
Same, ANCOVA on % Change from Baseline	1.78%	-14.38%	-16.10%	0.042
Same, ANCOVA on Rank of Change from Baseline				0.038
LoE imputation by mean of observed changes	-0.95	-1.95	-0.99	0.043
All imputation by mean of observed changes	-0.95	-2.01	-1.05	0.032
LOCF for all imputation	-0.72	-1.87	-1.15	0.018
Drop all missing values	-0.72	-2.11	-1.38	0.01
	n = 72	n = 79		

ANCOVA on change from baseline tested

Comment:
No analysis employing a fully non-parametric method, e.g., a rank sum test, was submitted. This was performed by CBER medical reviewer (without any covariate adjustments), and gave a p-value of 0.038, the same as reported by Allergan for use of ANCOVA on ranks.

The Physician Global Assessment rate of improvement was re-analyzed as requested, and showed 31.2% responders in placebo (n=82), 50.5% in BOTOX (n=88), for a difference of 19.4%, p=0.010.

Comment:
These analysis confirm the expectations from the initial review. On the prospective analytic plan for this primary endpoint the study failed to show statistically significant treatment effect. However, that appears to be, in an important manner, due to a flawed plan for analysis. While a parametric analytic method was selected, a missing data imputation method for LoE missing was used that creates a pool of outliers at the most extreme value observed in the study. This can be expected to seriously impair the performance of the ANCOVA method. Although this involves only a few subjects, it is sufficient in this small study. Other missing value imputation methods that do not create additional extreme outliers or analytic methods that are not as sensitive to the degree of outlier (rank based methods) provide a picture of the results that consistently indicates the study did show a marginally statistically significant treatment effect on this endpoint. All these analyses also consistently indicate that the size of the treatment effect is small, approximately 1 point on the CDSS scale. This is equivalent to an average of just 5 degrees change in head position in one plane in the treated group.